

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64878 A2

(51) International Patent Classification⁷: C12N 15/12,
C07K 14/47

TURNER, C., Alexander, Jr.; 67 Winter Wheat Place,
The Woodlands, TX 77381 (US).

(21) International Application Number: PCT/US01/06462

(74) Agents: ISHIMOTO, Lance, K. et al.; Lexicon Genetics
Incorporated, 4000 Research Forest Drive, The Woodlands,
TX 77381 (US).

(22) International Filing Date: 28 February 2001 (28.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/185,956 29 February 2000 (29.02.2000) US

(71) Applicant: LEXICON GENETICS INCORPORATED
[US/US]; 4000 Research Forest Drive, The Woodlands, TX
77381 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished
upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance
Notes on Codes and Abbreviations" appearing at the beginning
of each regular issue of the PCT Gazette.*

(72) Inventors: WALKE, D., Wade; 7507 Daneshill Drive,
Spring, TX 77389 (US). WILGANOWSKI, Nathaniel,
L.; Apartment 77, 9820 Memorial, Houston, TX 77024
(US). HU, Yi; 333 Holly Creek Court #203, The Wood-
lands, TX 77381 (US). KIEKE, James, Alvin; 9202
Restover Lane, Houston, TX 77064 (US). ZAMBROW-
ICZ, Brian; 18 Firethorne Place, The Woodlands, TX
77382 (US). REVELLI, Jean-Pierre; Apartment 115,
2255 Braeswood Park Drive, Houston, TX 77030 (US).

(54) Title: NOVEL HUMAN TRANSPORTER PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: Novel human polynucleotide and polypeptide sequences are disclosed that can be used in therapeutic, diagnostic,
and pharmacogenomic applications.

WO 01/64878 A2

NOVEL HUMAN TRANSPORTER PROTEINS AND
POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S.

5 Provisional Application Number 60/185,956 which was filed on
February 29, 2000 and is herein incorporated by reference in
its entirety.

1. INTRODUCTION

The present invention relates to the discovery,
10 identification, and characterization of novel human
polynucleotides encoding proteins that share sequence
similarity with mammalian transporter proteins. The invention
encompasses the described polynucleotides, host cell expression
systems, the encoded proteins, fusion proteins, polypeptides
15 and peptides, antibodies to the encoded proteins and peptides,
and genetically engineered animals that either lack or over
express the disclosed polynucleotides, antagonists and agonists
of the proteins, and other compounds that modulate the
expression or activity of the proteins encoded by the disclosed
20 polynucleotides that can be used for diagnosis, drug screening,
clinical trial monitoring, and treatment of diseases and
disorders.

2. BACKGROUND OF THE INVENTION

Transporter proteins are integral membrane proteins that
25 mediate or facilitate the passage of materials across the lipid
bilayer. Given that the transport of materials across the
membrane can play an important physiological role, transporter
proteins are good drug targets. Additionally, one of the
mechanisms of drug resistance involves diseased cells using
30 cellular transporter systems to export chemotherapeutic agents
from the cell. Such mechanisms are particularly relevant to
cells manifesting resistance to a multiplicity of drugs.

3. SUMMARY OF THE INVENTION

The present invention relates to the discovery, identification, and characterization of nucleotides that encode novel human proteins, and the corresponding amino acid sequences of these proteins. The novel human proteins (NHPs) described for the first time herein share structural similarity with mammalian ion transporters, calcium transporters (particularly calcium transporting ATPases), sulfate transporters, and zinc transporters.

The novel human nucleic acid sequences described herein, encode alternative proteins/open reading frames (ORFs) of 1,177 and 374 amino acids in length (calcium-transporting ATPase, SEQ ID NOS: 2 and 4), 970 (sulfate transporter, SEQ ID NO:7), and 507 (zinc transporter, SEQ ID NO:10) amino acids in length.

The invention also encompasses agonists and antagonists of the described NHPs, including small molecules, large molecules, mutant NHPs, or portions thereof, that compete with native NHP, peptides, and antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NHPs

(e.g., antisense and ribozyme molecules, and gene or regulatory sequence replacement constructs) or to enhance the expression of the described NHP polynucleotides (e.g., expression constructs that place the described polynucleotide under the control of a strong promoter system), and transgenic animals

that express a NHP transgene, or "knock-outs" (which can be conditional) that do not express a functional NHP. Knock-out mice can be produced in several ways, one of which involves the use of mouse embryonic stem cells ("ES cells") lines that contain gene trap mutations in a murine homolog of at least one

of the described NHPs. When the unique NHP sequences described in SEQ ID NOS:1-11 are "knocked-out" they provide a method of identifying phenotypic expression of the particular gene as well as a method of assigning

function to previously unknown genes. Additionally, the unique NHP sequences described in SEQ ID NOS:1-11 are useful for the identification of coding sequence and the mapping a unique gene to a particular chromosome.

5 Further, the present invention also relates to processes for identifying compounds that modulate, i.e., act as agonists or antagonists, of NHP expression and/or NHP activity that utilize purified preparations of the described NHPs and/or NHP product, or cells expressing the same. Such compounds can be
10 used as therapeutic agents for the treatment of any of a wide variety of symptoms associated with biological disorders or imbalances.

4. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

15 The Sequence Listing provides the sequences of the described NHP ORFs that encode the described NHP amino acid sequences. SEQ ID NOS 5, 8, and 11 describe nucleotides encoding NHP ORFs along with regions of flanking sequence.

5. DETAILED DESCRIPTION OF THE INVENTION

20 The NHPs described for the first time herein are novel proteins that may be expressed in, *inter alia*, human cell lines, fetal brain, pituitary, cerebellum, thymus, spleen, lymph node, bone marrow, trachea, kidney, fetal liver, liver,
25 prostate, testis, thyroid, adrenal gland, salivary gland, stomach, small intestine, colon, adipose, rectum, pericardium, bone marrow, placenta, and gene trapped human cells. More particularly, the NHP that is similar to sulfate transporters (and the down-regulated in adenoma, or DRA, gene) is
30 predominantly found in bone marrow and testis, and the zinc transporter-like NHP can be found expressed in the placenta.

The present invention encompasses the nucleotides presented in the Sequence Listing, host cells expressing such

nucleotides, the expression products of such nucleotides, and:
(a) nucleotides that encode mammalian homologs of the described polynucleotides, including the specifically described NHPs, and the NHP products; (b) nucleotides that encode one or more
5 portions of the NHPs that correspond to functional domains, and the polypeptide products specified by such nucleotide sequences, including but not limited to the novel regions of any active domain(s); (c) isolated nucleotides that encode mutant versions, engineered or naturally occurring, of the
10 described NHPs in which all or a part of at least one domain is deleted or altered, and the polypeptide products specified by such nucleotide sequences, including but not limited to soluble proteins and peptides in which all or a portion of the signal (or hydrophobic transmembrane) sequence is deleted; (d)
15 nucleotides that encode chimeric fusion proteins containing all or a portion of a coding region of an NHP, or one of its domains (e.g., a receptor or ligand binding domain, accessory protein/self-association domain, etc.) fused to another peptide or polypeptide; or (e) therapeutic or diagnostic derivatives of
20 the described polynucleotides such as oligonucleotides, antisense polynucleotides, ribozymes, dsRNA, or gene therapy constructs comprising a sequence first disclosed in the Sequence Listing. As discussed above, the present invention includes: (a) the human DNA sequences presented in the Sequence
25 Listing (and vectors comprising the same) and additionally contemplates any nucleotide sequence encoding a contiguous NHP open reading frame (ORF) that hybridizes to a complement of a DNA sequence presented in the Sequence Listing under highly stringent conditions, e.g., hybridization to filter-bound DNA
30 in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons,

Inc., New York, at p. 2.10.3) and encodes a functionally equivalent gene product. Additionally contemplated are any nucleotide sequences that hybridize to the complement of a DNA sequence that encodes and expresses an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, e.g., washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*), yet still encodes a functionally equivalent NHP product. Functional equivalents of a NHP include naturally occurring NHPs present in other species and mutant NHPs whether naturally occurring or engineered (by site directed mutagenesis, gene shuffling, directed evolution as described in, for example, U.S. Patent No. 5,837,458). The invention also includes degenerate nucleic acid variants of the disclosed NHP polynucleotide sequences.

Additionally contemplated are polynucleotides encoding NHP ORFs, or their functional equivalents, encoded by polynucleotide sequences that are about 99, 95, 90, or about 85 percent similar or identical to corresponding regions of the nucleotide sequences of the Sequence Listing (as measured by BLAST sequence comparison analysis using, for example, the GCG sequence analysis package using standard default settings).

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore the complements of, the described NHP nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described above. In instances where the nucleic acid molecules are deoxyoligonucleotides ("DNA oligos"), such molecules are generally about 16 to about 100 bases long, or about 20 to about 80, or about 34 to about 45 bases long, or any variation or combination of sizes represented therein that incorporate a contiguous region of sequence first disclosed in the Sequence Listing. Such oligonucleotides can be used in conjunction with the polymerase

chain reaction (PCR) to screen libraries, isolate clones, and prepare cloning and sequencing templates, etc.

Alternatively, such NHP oligonucleotides can be used as hybridization probes for screening libraries, and assessing gene expression patterns (particularly using a micro array or high-throughput "chip" format). Additionally, a series of the described NHP oligonucleotide sequences, or the complements thereof, can be used to represent all or a portion of the described NHP sequences. An oligonucleotide or polynucleotide sequence first disclosed in at least a portion of one or more of the sequences of SEQ ID NOS: 1-11 can be used as a hybridization probe in conjunction with a solid support matrix/substrate (resins, beads, membranes, plastics, polymers, metal or metallized substrates, crystalline or polycrystalline substrates, etc.). Of particular note are spatially addressable arrays (i.e., gene chips, microtiter plates, etc.) of oligonucleotides and polynucleotides, or corresponding oligopeptides and polypeptides, wherein at least one of the biopolymers present on the spatially addressable array comprises an oligonucleotide or polynucleotide sequence first disclosed in at least one of the sequences of SEQ ID NOS: 1-11, or an amino acid sequence encoded thereby. Methods for attaching biopolymers to, or synthesizing biopolymers on, solid support matrices, and conducting binding studies thereon are disclosed in, *inter alia*, U.S. Patent Nos. 5,700,637, 5,556,752, 5,744,305, 4,631,211, 5,445,934, 5,252,743, 4,713,326, 5,424,186, and 4,689,405 the disclosures of which are herein incorporated by reference in their entirety.

Addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-11 can be used to identify and characterize the temporal and tissue specific expression of a gene. These addressable arrays incorporate oligonucleotide sequences of sufficient length to confer the required specificity, yet be

within the limitations of the production technology. The length of these probes is within a range of between about 8 to about 2000 nucleotides. Preferably the probes consist of 60 nucleotides and more preferably 25 nucleotides from the sequences first disclosed in SEQ ID NOS:1-11.

For example, a series of the described oligonucleotide sequences, or the complements thereof, can be used in chip format to represent all or a portion of the described sequences. The oligonucleotides, typically between about 16 to about 40 (or any whole number within the stated range) nucleotides in length can partially overlap each other and/or the sequence may be represented using oligonucleotides that do not overlap. Accordingly, the described polynucleotide sequences shall typically comprise at least about two or three distinct oligonucleotide sequences of at least about 8 nucleotides in length that are each first disclosed in the described Sequence Listing. Such oligonucleotide sequences can begin at any nucleotide present within a sequence in the Sequence Listing and proceed in either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense orientation.

Microarray-based analysis allows the discovery of broad patterns of genetic activity, providing new understanding of gene functions and generating novel and unexpected insight into transcriptional processes and biological mechanisms. The use of addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-11 provides detailed information about transcriptional changes involved in a specific pathway, potentially leading to the identification of novel components or gene functions that manifest themselves as novel phenotypes.

Probes consisting of sequences first disclosed in SEQ ID NOS:1-11 can also be used in the identification, selection and validation of novel molecular targets for drug discovery. The

use of these unique sequences permits the direct confirmation of drug targets and recognition of drug dependent changes in gene expression that are modulated through pathways distinct from the drugs intended target. These unique sequences
5 therefore also have utility in defining and monitoring both drug action and toxicity.

As an example of utility, the sequences first disclosed in SEQ ID NOS:1-11 can be utilized in microarrays or other assay formats, to screen collections of genetic material from
10 patients who have a particular medical condition. These investigations can also be carried out using the sequences first disclosed in SEQ ID NOS:1-11 *in silico* and by comparing previously collected genetic databases and the disclosed sequences using computer software known to those in the art.

15 Thus the sequences first disclosed in SEQ ID NOS:1-11 can be used to identify mutations associated with a particular disease and also as a diagnostic or prognostic assay.

Although the presently described sequences have been specifically described using nucleotide sequence, it should be
20 appreciated that each of the sequences can uniquely be described using any of a wide variety of additional structural attributes, or combinations thereof. For example, a given sequence can be described by the net composition of the nucleotides present within a given region of the sequence in
25 conjunction with the presence of one or more specific oligonucleotide sequence(s) first disclosed in the SEQ ID NOS: 1-11. Alternatively, a restriction map specifying the relative positions of restriction endonuclease digestion sites, or various palindromic or other specific oligonucleotide sequences
30 can be used to structurally describe a given sequence. Such restriction maps, which are typically generated by widely available computer programs (e.g., the University of Wisconsin GCG sequence analysis package, SEQUENCHER 3.0, Gene Codes

Corp., Ann Arbor, MI, etc.), can optionally be used in conjunction with one or more discrete nucleotide sequence(s) present in the sequence that can be described by the relative position of the sequence relative to one or more additional
5 sequence(s) or one or more restriction sites present in the disclosed sequence.

For oligonucleotide probes, highly stringent conditions may refer, e.g., to washing in 6xSSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C
10 (for 20-base oligos), and 60°C (for 23-base oligos). These nucleic acid molecules may encode or act as NHP gene antisense molecules, useful, for example, in NHP gene regulation (for and/or as antisense primers in amplification reactions of NHP gene nucleic acid sequences). With respect to NHP gene
15 regulation, such techniques can be used to regulate biological functions. Further, such sequences may be used as part of ribozyme and/or triple helix sequences that are also useful for NHP gene regulation.

Inhibitory antisense or double stranded oligonucleotides
20 can additionally comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-
25 2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine,
30 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine,

2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil,
4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid
methylester, uracil-5-oxyacetic acid (v), 5-methyl-
2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w,
5 and 2,6-diaminopurine.

The antisense oligonucleotide can also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

10 In yet another embodiment, the antisense oligonucleotide will comprise at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl
15 phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an α -anomeric oligonucleotide. An α -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the
20 strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide is a 2'-O-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330). Alternatively, double stranded
25 RNA can be used to disrupt the expression and function of a targeted NHP.

Oligonucleotides of the invention can be synthesized by standard methods known in the art, e.g. by use of an automated DNA synthesizer (such as are commercially available from
30 Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides can be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), and methylphosphonate oligonucleotides can be prepared by use of

controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

Low stringency conditions are well known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual (and periodic updates thereof), Cold Springs Harbor Press, N.Y.; and Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.

Alternatively, suitably labeled NHP nucleotide probes can be used to screen a human genomic library using appropriately stringent conditions or by PCR. The identification and characterization of human genomic clones is helpful for identifying polymorphisms (including, but not limited to, nucleotide repeats, microsatellite alleles, single nucleotide polymorphisms, or coding single nucleotide polymorphisms), determining the genomic structure of a given locus/allele, and designing diagnostic tests. For example, sequences derived from regions adjacent to the intron/exon boundaries of the human gene can be used to design primers for use in amplification assays to detect mutations within the exons, introns, splice sites (e.g., splice acceptor and/or donor sites), etc., that can be used in diagnostics and pharmacogenomics.

Further, a NHP gene homolog can be isolated from nucleic acid from an organism of interest by performing PCR using two degenerate or "wobble" oligonucleotide primer pools designed on the basis of amino acid sequences within the NHP products disclosed herein. The template for the reaction may be total RNA, mRNA, and/or cDNA obtained by reverse transcription of

mRNA prepared from human or non-human cell lines or tissue known or suspected to express an allele of a NHP gene.

The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the
5 desired NHP gene. The PCR fragment can then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment can be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment can be used to isolate genomic clones via
10 the screening of a genomic library.

PCR technology can also be used to isolate full length cDNA sequences. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source (i.e., one known, or suspected, to express a NHP gene).
15 A reverse transcription (RT) reaction can be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" using a standard terminal transferase reaction, the hybrid may
20 be digested with RNase H, and second strand synthesis may then be primed with a complementary primer. Thus, cDNA sequences upstream of the amplified fragment can be isolated. For a review of cloning strategies that can be used, see e.g., Sambrook et al., 1989, *supra*.

25 A cDNA encoding a mutant NHP gene can be isolated, for example, by using PCR. In this case, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known or suspected to be expressed in an individual putatively carrying a mutant NHP allele, and by
30 extending the new strand with reverse transcriptase. The second strand of the cDNA is then synthesized using an oligonucleotide that hybridizes specifically to the 5' end of the normal gene. Using these two primers, the product is then

amplified via PCR, optionally cloned into a suitable vector, and subjected to DNA sequence analysis through methods well known to those of skill in the art. By comparing the DNA sequence of the mutant NHP allele to that of a corresponding normal NHP allele, the mutation(s) responsible for the loss or alteration of function of the mutant NHP gene product can be ascertained.

Alternatively, a genomic library can be constructed using DNA obtained from an individual suspected of or known to carry a mutant NHP allele (e.g., a person manifesting a NHP-associated phenotype such as, for example, obesity, high blood pressure, connective tissue disorders, infertility, etc.), or a cDNA library can be constructed using RNA from a tissue known, or suspected, to express a mutant NHP allele. A normal NHP gene, or any suitable fragment thereof, can then be labeled and used as a probe to identify the corresponding mutant NHP allele in such libraries. Clones containing mutant NHP gene sequences can then be purified and subjected to sequence analysis according to methods well known to those skilled in the art.

Additionally, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known, or suspected, to express a mutant NHP allele in an individual suspected of or known to carry such a mutant allele. In this manner, gene products made by the putatively mutant tissue can be expressed and screened using standard antibody screening techniques in conjunction with antibodies raised against a normal NHP product, as described below. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, NY).

Additionally, screening can be accomplished by screening with labeled NHP fusion proteins, such as, for example, alkaline phosphatase-NHP or NHP-alkaline phosphatase fusion

proteins. In cases where a NHP mutation results in an expressed gene product with altered function (e.g., as a result of a missense or a frameshift mutation), polyclonal antibodies to a NHP are likely to cross-react with a corresponding mutant
5 NHP gene product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis according to methods well known in the art.

The invention also encompasses (a) DNA vectors that contain any of the foregoing NHP coding sequences and/or their
10 complements (i.e., antisense); (b) DNA expression vectors that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences (for example, baculo virus as described in U.S. Patent No. 5,869,336 herein incorporated
15 by reference); (c) genetically engineered host cells that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell; and (d) genetically engineered host cells that express an endogenous
20 NHP gene under the control of an exogenously introduced regulatory element (i.e., gene activation). As used herein, regulatory elements include, but are not limited to, inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and
25 regulate expression. Such regulatory elements include but are not limited to the cytomegalovirus (hCMV) immediate early gene, regulatable, viral elements (particularly retroviral LTR promoters), the early or late promoters of SV40 adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the
30 major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast α -mating factors.

The present invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists and agonists of the NHP, as well as compounds or nucleotide constructs that inhibit expression of a NHP gene
5 (transcription factor inhibitors, antisense and ribozyme molecules, or gene or regulatory sequence replacement constructs), or promote the expression of a NHP (e.g., expression constructs in which NHP coding sequences are operatively associated with expression control elements such as
10 promoters, promoter/enhancers, etc.).

The NHPs or NHP peptides, NHP fusion proteins, NHP nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant NHPs or inappropriately expressed NHPs for the diagnosis of disease. The NHP proteins
15 or peptides, NHP fusion proteins, NHP nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for screening for drugs (or high throughput screening of combinatorial libraries) effective in the treatment of the
20 symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. The use of engineered host cells and/or animals may offer an advantage in that such systems allow not only for the identification of compounds that bind to the endogenous receptor for an NHP, but can also
25 identify compounds that trigger NHP-mediated activities or pathways.

Finally, the NHP products can be used as therapeutics. For example, soluble derivatives such as NHP peptides/domains corresponding to NHPs, NHP fusion protein products (especially
30 NHP-Ig fusion proteins, i.e., fusions of a NHP, or a domain of a NHP, to an IgFc), NHP antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists or agonists (including compounds that modulate or act on downstream targets

in a NHP-mediated pathway) can be used to directly treat diseases or disorders. For instance, the administration of an effective amount of soluble NHP, or a NHP-IgFc fusion protein or an anti-idiotypic antibody (or its Fab) that mimics the NHP could activate or effectively antagonize the endogenous NHP receptor. Nucleotide constructs encoding such NHP products can be used to genetically engineer host cells to express such products *in vivo*; these genetically engineered cells function as "bioreactors" in the body delivering a continuous supply of a NHP, a NHP peptide, or a NHP fusion protein to the body. Nucleotide constructs encoding functional NHPs, mutant NHPs, as well as antisense and ribozyme molecules can also be used in "gene therapy" approaches for the modulation of NHP expression. Thus, the invention also encompasses pharmaceutical formulations and methods for treating biological disorders.

Various aspects of the invention are described in greater detail in the subsections below.

5.1 THE NHP SEQUENCES

The cDNA sequences and the corresponding deduced amino acid sequences of the described NHPs are presented in the Sequence Listing. The NHP nucleotides were obtained from clustered human gene trapped sequences, testis and mammary transcript RACE products, ESTs, and human brain, testis, trachea, pituitary, thymus, and mammary gland cDNA libraries (Edge Biosystems, Gaithersburg, MD).

SEQ ID NOS:1-5 describe sequences that are similar to eucaryotic ATP-driven ion pumps such as calcium transporting ATPases, and which can be found expressed in a variety of human cells and tissues. The described sequences were assembled using gene trapped sequences and clones isolated from human kidney, lymph node, and thymus cDNA libraries (Edge Biosystems, Gaithersburg, MD).

SEQ ID NOS:6-8 describe sequences that are similar to, *inter alia*, sulfate transporter and cotransporter proteins, and can be found expressed in human bone marrow and testis. Several polymorphisms were found in this NHP including, but not limited to, possible A-to-G transitions at nucleotide positions corresponding to nucleotides 589, 692, 917, 1,164, and 2,390 of, for example SEQ ID NO:8 which be silent or can result in the met corresponding to amino acid position 73 of SEQ ID NO:7 converting to a val (e.g., met 73 converting to val 73), val 148 converting to ile, asn 230 converting to lys, ile 562 converting to val. An additional C-to-T transition was identified that converts ala 777 to val. SEQ ID NOS:6-8 can be expressed in bone marrow and predominantly in testis cells. These NHPs were assembled from gene trapped sequences and clones from a human testis cDNA library (Edge Biosystems, Gaithersburg, MD).

SEQ ID NOS:9-11 describe sequences that are similar to zinc transporters and vesicular transporters, can be found expressed in, *inter alia*, placenta and adrenal gland, and these NHP sequences were assembled using gene trapped sequences and clones from human adrenal and placenta cDNA libraries (Edge Biosystems, Gaithersburg, MD).

Transporters and transporter related multidrug resistance (MDR) sequences, as well as uses and applications that are germane to the described NHPs, are described in U.S. Patents Nos. 5,198,344 and 5,866,699 which are herein incorporated by reference in their entirety.

5.2 NHPS AND NHP POLYPEPTIDES

NHPs, polypeptides, peptide fragments, mutated, truncated, or deleted forms of the NHPs, and/or NHP fusion proteins can be prepared for a variety of uses. These uses include but are not limited to the generation of antibodies, as reagents in

diagnostic assays, the identification of other cellular gene products related to a NHP, as reagents in assays for screening for compounds that can be as pharmaceutical reagents useful in the therapeutic treatment of mental, biological, or medical disorders and diseases. Given the similarity information and expression data, the described NHPs can be targeted (by drugs, oligos, antibodies, etc,) in order to treat disease, or to therapeutically augment the efficacy of, for example, chemotherapeutic agents used in the treatment of breast or prostate cancer.

The Sequence Listing discloses the amino acid sequences encoded by the described NHP polynucleotides. The NHPs typically display have initiator methionines in DNA sequence contexts consistent with a translation initiation site.

The NHP amino acid sequences of the invention include the amino acid sequence presented in the Sequence Listing as well as analogues and derivatives thereof. Further, corresponding NHP homologues from other species are encompassed by the invention. In fact, any NHP protein encoded by the NHP nucleotide sequences described above are within the scope of the invention, as are any novel polynucleotide sequences encoding all or any novel portion of an amino acid sequence presented in the Sequence Listing. The degenerate nature of the genetic code is well known, and, accordingly, each amino acid presented in the Sequence Listing, is generically representative of the well known nucleic acid "triplet" codon, or in many cases codons, that can encode the amino acid. As such, as contemplated herein, the amino acid sequences presented in the Sequence Listing, when taken together with the genetic code (see, for example, Table 4-1 at page 109 of "Molecular Cell Biology", 1986, J. Darnell et al. eds., Scientific American Books, New York, NY, herein incorporated by reference) are generically representative of all the various

permutations and combinations of nucleic acid sequences that can encode such amino acid sequences.

The invention also encompasses proteins that are functionally equivalent to the NHPs encoded by the presently described nucleotide sequences as judged by any of a number of criteria, including, but not limited to, the ability to bind and cleave a substrate of a NHP, or the ability to effect an identical or complementary downstream pathway, or a change in cellular metabolism (e.g., proteolytic activity, ion flux, tyrosine phosphorylation, etc.). Such functionally equivalent NHP proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the NHP nucleotide sequences described above, but which result in a silent change, thus producing a functionally equivalent gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

A variety of host-expression vector systems can be used to express the NHP nucleotide sequences of the invention. Where, as in the present instance, the NHP peptide or polypeptide is thought to be membrane protein, the hydrophobic regions of the protein can be excised and the resulting soluble peptide or polypeptide can be recovered from the culture media. Such expression systems also encompass engineered host cells that express a NHP, or functional equivalent, *in situ*. Purification

or enrichment of a NHP from such expression systems can be accomplished using appropriate detergents and lipid micelles and methods well known to those skilled in the art. However, such engineered host cells themselves may be used in situations
5 where it is important not only to retain the structural and functional characteristics of the NHP, but to assess biological activity, e.g., in drug screening assays.

The expression systems that may be used for purposes of the invention include but are not limited to microorganisms
10 such as bacteria (e.g., *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing NHP nucleotide sequences; yeast (e.g., *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing NHP nucleotide sequences;
15 insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing NHP sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors
20 (e.g., Ti plasmid) containing NHP nucleotide sequences; or mammalian cell systems (e.g., COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the
25 adenovirus late promoter; the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the NHP product being expressed. For example, when a large quantity of such a protein is to be produced for the generation
30 of pharmaceutical compositions of or containing NHP, or for raising antibodies to a NHP, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not

limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, EMBO J. 2:1791), in which a NHP coding sequence may be ligated individually into the vector in frame with the *lacZ* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Pharmacia or American Type Culture Collection) can also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The PGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

In an insect system, *Autographa californica* nuclear polyhidrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. A NHP coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of NHP coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted sequence is expressed (e.g., see Smith et al., 1983, J. Virol. 46:584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the NHP nucleotide

sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a NHP product in infected hosts (e.g., See Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted NHP nucleotide sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire NHP gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of a NHP coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (See Bittner et al., 1987, Methods in Enzymol. 153:516-544).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and

specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed.

5 To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and in particular, human cell lines.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the NHP sequences described above can be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the NHP product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the NHP product.

30 A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, et al., 1977, Cell 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc.

Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy, et al., 1980, Cell 22:817) genes can be employed in tk⁻, hgp⁻ or ap⁻ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler, et al., 1980, Natl. Acad. Sci. USA 77:3567; O'Hare, et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, et al., 1981, J. Mol. Biol. 150:1); and hyg^r, which confers resistance to hygromycin (Santerre, et al., 1984, Gene 30:147).

Alternatively, any fusion protein can be readily purified by utilizing an antibody specific for the fusion protein being expressed. For example, a system described by Janknecht et al. allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, et al., 1991, Proc. Natl. Acad. Sci. USA 88:8972-8976). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni²⁺-nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Also encompassed by the present invention are fusion proteins that direct the NHP to a target organ and/or facilitate transport across the membrane into the cytosol. Conjugation of NHPs to antibody molecules or their Fab fragments could be used to target cells bearing a particular epitope. Attaching the appropriate signal sequence to the NHP

would also transport the NHP to the desired location within the cell. Alternatively targeting of NHP or its nucleic acid sequence might be achieved using liposome or lipid complex based delivery systems. Such technologies are described in

5 Liposomes: A Practical Approach, New, RRC ed., Oxford University Press, New York and in U.S. Patents Nos. 4,594,595, 5,459,127, 5,948,767 and 6,110,490 and their respective disclosures which are herein incorporated by reference in their entirety.

Additionally embodied are novel protein constructs engineered

10 in such a way that they facilitate transport of the NHP to the target site or desired organ. This goal may be achieved by coupling of the NHP to a cytokine or other ligand that provides targeting specificity, and/or to a protein transducing domain (see generally U.S. applications Ser. No. 60/111,701 and

15 60/056,713, both of which are herein incorporated by reference, for examples of such transducing sequences) to facilitate passage across cellular membranes if needed and can optionally be engineered to include nuclear localization sequences when desired.

20

5.3 ANTIBODIES TO NHP PRODUCTS

Antibodies that specifically recognize one or more epitopes of a NHP, or epitopes of conserved variants of a NHP, or peptide fragments of a NHP are also encompassed by the

25 invention. Such antibodies include but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-

30 binding fragments of any of the above.

The antibodies of the invention may be used, for example, in the detection of NHP in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic

technique whereby patients may be tested for abnormal amounts of NHP. Such antibodies may also be utilized in conjunction with, for example, compound screening schemes for the evaluation of the effect of test compounds on expression and/or activity of a NHP gene product. Additionally, such antibodies can be used in conjunction gene therapy to, for example, evaluate the normal and/or engineered NHP-expressing cells prior to their introduction into the patient. Such antibodies may additionally be used as a method for the inhibition of abnormal NHP activity. Thus, such antibodies may, therefore, be utilized as part of treatment methods.

For the production of antibodies, various host animals may be immunized by injection with a NHP, an NHP peptide (e.g., one corresponding to a functional domain of an NHP), truncated NHP polypeptides (NHP in which one or more domains have been deleted), functional equivalents of the NHP or mutated variant of the NHP. Such host animals may include but are not limited to pigs, rabbits, mice, goats, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's adjuvant (complete and incomplete), mineral salts such as aluminum hydroxide or aluminum phosphate, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Alternatively, the immune response could be enhanced by combination and or coupling with molecules such as keyhole limpet hemocyanin, tetanus toxoid, diphtheria toxoid, ovalbumin, cholera toxin or fragments thereof.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be obtained by any

technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, Nature 256:495-497; and U.S. Patent No.

5 4,376,110), the human B-cell hybridoma technique (Kosbor et al., 1983, Immunology Today 4:72; Cole et al., 1983, Proc. Natl. Acad. Sci. USA 80:2026-2030), and the EBV-hybridoma technique (Cole et al., 1985, Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may
10 be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

15 In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci., 81:6851-6855; Neuberger et al., 1984, Nature, 312:604-608; Takeda et al., 1985, Nature, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen
20 specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin
25 constant region. Such technologies are described in U.S. Patents Nos. 6,075,181 and 5,877,397 and their respective disclosures which are herein incorporated by reference in their entirety. Also encompassed by the present invention is the use
30 of fully humanized monoclonal antibodies as described in US Patent No. 6,150,584 and respective disclosures which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; Bird, 1988, Science 242:423-426; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward et al., 1989, Nature 334:544-546) can be adapted to produce single chain antibodies against NHP gene products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include, but are not limited to: the F(ab')₂ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., 1989, Science, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to a NHP can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" a given NHP, using techniques well known to those skilled in the art. (See, e.g., Greenspan & Bona, 1993, FASEB J 7(5):437-444; and Nissinoff, 1991, J. Immunol. 147(8):2429-2438). For example antibodies which bind to a NHP domain and competitively inhibit the binding of NHP to its cognate receptor can be used to generate anti-idiotypes that "mimic" the NHP and, therefore, bind and activate or neutralize a receptor. Such anti-idiotypic antibodies or Fab fragments of such anti-idiotypes can be used in therapeutic regimens involving a NHP mediated pathway.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within

the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall
5 within the scope of the appended claims. All cited publications, patents, and patent applications are herein incorporated by reference in their entirety.

10

15

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising at least 24 contiguous bases of nucleotide sequence first disclosed in SEQ ID NO: 1.
5
2. An isolated nucleic acid molecule comprising a nucleotide sequence that:
 - (a) encodes the amino acid sequence shown in SEQ ID NO: 2; and
10
 - (b) hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:1 or the complement thereof.
- 15 3. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:2.
- 20 4. An isolated nucleic acid molecule comprising a nucleotide sequence that:
 - (a) encodes the amino acid sequence shown in SEQ ID NO:7; and
 - (b) hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:6 or the
25 complement thereof.
- 30 5. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:7.
6. An isolated nucleic acid molecule comprising a nucleotide sequence that:

- (a) encodes the amino acid sequence shown in SEQ ID NO:10; and
- (b) hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:9 or the complement thereof.

5

7.. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:10.

10

SEQUENCE LISTING

<110> LEXICON GENETICS INCORPORATED

<120> Novel Human Transporter Proteins and Polynucleotides Encoding the Same

<130> LEX-0141-PCT

<150> US 60/185,956

<151> 2000-02-29

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 3534

<212> DNA

<213> homo sapiens

<400> 1

atgtggcgct	ggatccggca	gcagctgggt	tttgaccac	cacatcagag	tgacacaaga	60
accatctacg	tagccaacag	gtttcctcag	aatggccttt	acacacctca	gaaatttata	120
gataacagga	tcatttcac	taagtacact	gtgtggaatt	ttgttccaaa	aaattttatt	180
gaacagttca	gaagagtggc	aaacttttat	tttcttatta	tatttttggg	tcagcttatg	240
attgatacac	ctaccagtcc	agttaccagt	ggacttccat	tattctttgt	gataacagta	300
actgccataa	agcaggggata	tgaagattgg	ttacggcata	actcagataa	tgaagtaaat	360
ggagctcctg	tttatgttgt	tcgaagtggg	ggccttgtaa	aaactagatc	aaaaaacatt	420
cgggtgggtg	atattgttcg	aatagccaaa	gatgaaattt	ttcctgcaga	cttggtgctt	480
ctgtcctcag	atcgactgga	tggttcctgt	cacgttacaa	ctgctagttt	ggacggagaa	540
actaacctga	agacacatgt	ggcagttcca	gaaacagcat	tattacaac	agttgccaat	600
ttggacactc	tagtagctgt	aatagaatgc	cagcaaccag	aagcagactt	atacagattc	660
atgggacgaa	tgatcataac	ccaacaaatg	gaagaaattg	taagacctct	ggggccggag	720
agtctcctgc	ttcgtggagc	cagattaaaa	aacacaaaag	aaatttttgg	tggtgcggta	780
tacactggaa	tgaaactaa	gatggcatta	aattacaaga	gcaaatcaca	gaaacgatct	840
gcagttagaaa	agtcaatgaa	tacatttttg	ataattttat	tagtaattct	tatatctgaa	900
gctgtcatca	gcactatctt	gaagtataca	tggcaagctg	aagaaaaatg	ggatgaacct	960
tgggtataacc	aaaaaacaga	acatcaaaga	aatagcagta	agattctgag	atttatttca	1020
gacttccttg	cttttttggg	tctctacaat	ttcatcattc	caatttcatt	atatgtgaca	1080
gtcgaaatgc	agaaatttct	tggatcattt	tttattggct	gggatcttga	tctgtatcat	1140
gaagaatcag	atcagaaagc	tcaagtcaat	acttccgac	tgaatgaaga	gcttggacag	1200
gtagagtacg	tggttacaga	taaaactggg	acactgacag	aaaatgagat	gcagtttcgg	1260
gaatgttcaa	ttaatggcat	gaaataccaa	gaaattaatg	gtagacttgt	acccgaagga	1320
ccaacaccag	actcttcaga	aggaaactta	tcttatctta	gtagtttatc	ccatcttaac	1380
aacttatccc	atcttacaac	cagttcctct	ttcagaacca	gtcctgaaaa	tgaaactgaa	1440
ctaattaaag	aacatgatct	cttctttaaa	gcagtcagtc	tctgtcacac	tgtacagatt	1500
agcaatgttc	aaactgactg	cactgggtgat	ggtccctggc	aatccaacct	ggcaccatcg	1560
cagttggagt	actatgcac	ttcaccagat	gaaaaggctc	tagtagaagc	tgctgcaagg	1620
attggatttg	tgtttattgg	caattctgaa	gaaactatgg	agggttaaac	tcttggaata	1680
ctggaacggt	acaaactgct	tcatattctg	gaatttgatt	cagatcgtag	gagaatgagt	1740
gtaattgttc	aggcaccttc	aggtgagaag	ttattatttg	ctaaaggagc	tgagtcacat	1800
attctcccta	aatgtatagg	tgagaaata	gaaaaaacca	gaattcatgt	agatgaattt	1860
gctttgaaag	ggctaagaac	tctgtgtata	gcatatagaa	aatttacatc	aaaagagtat	1920
gaggaaatag	ataaacgcat	atttgaagcc	aggactgcct	tgcagcagcg	ggaagagaaa	1980

```

ttggcagctg ttttccagtt catagagaaa gacctgatat tacttggagc cacagcagta 2040
gaagacagac tacaagataa agttcgagaa actattgaag cattgagaat ggctggatc 2100
aaagtatggg tacttactgg ggataaacat gaaacagctg ttagtgtag tttatcatgt 2160
ggccattttc atagaacat gaacatcctt gaacttataa accagaaatc agacagcgag 2220
tgtgctgaac aattgaggca gcttgccaga agaattacag aggatcatgt gattcagcat 2280
gggctgtag tggatgggac cagcctatct cttgcactca gggagcatga aaaactattt 2340
atggaagttt gcagaaattg ttcagctgta ttatgctgtc gtatggctcc actgcagaaa 2400
gcaaaagtaa taagactaat aaaaatatca cctgagaaac ctataacatt ggctgttggt 2460
gatggtgcta atgacgtaag catgatacaa gaagcccatg ttggcatagg aatcatgggt 2520
aaagaaggaa gacaggctgc aagaaacagt gactatgcaa tagccagatt taagttcctc 2580
tccaaattgc tttttgttca tggctatttt tattatatta gaatagctac ccttgtacag 2640
tatttttttt ataagaatgt gtgctttatc acacccagct ttttatatca gttctactgt 2700
ttgttttctc agcaaacatt gtatgacagc gtgtacctga ctttatacaa tatttgtttt 2760
acttccctac ctattctgat atatagtctt ttggaacagc atgtagacct tcatgtgtta 2820
caaaataagc ccacccttta tgcagacatt agtaaaaacc gcctcttaag tattaaaaca 2880
tttctttatt ggaccatcct gggcttcagt catgccttta ttttcttttt tggatcctat 2940
ttactaatag ggaaagatac atctctgctt ggaaatggcc agatgttygg aaactggaca 3000
tttggcactt tggctctcac agtcatggtt attacagtca cagtaaagat ggctctggaa 3060
actcattttt ggacttggat caaccatctc gttacctggg gatctattat attttatttt 3120
gtattttcct tgttttatgg agggattctc tggccatttt tgggctccca gaatatgtat 3180
tttgtgttta ttcagctcct gtcaaagtgt tctgcttggt ttgccataat cctcatgggt 3240
gttacctgtc tatttcttga tatcataaag aaggcttttg accgacacct ccaccctaca 3300
agtactgaaa aggcacagct tactgaaaca aatgcaggta tcaagtgtt ggactccatg 3360
tgctgtttcc cggaaggaga agcagcgtgt gcatctgttg gaagaatgct ggaacgagtt 3420
ataggaagat gtagtccaac ccacatcagc agatcatgga gtgcatcgga tcctttctat 3480
accaacgaca ggagcatctt gactctctcc acaatggact catctacttg ttaa 3534

```

<210> 2

<211> 1177

<212> PRT

<213> homo sapiens

<400> 2

```

Met Trp Arg Trp Ile Arg Gln Gln Leu Gly Phe Asp Pro Pro His Gln
  1             5             10             15
Ser Asp Thr Arg Thr Ile Tyr Val Ala Asn Arg Phe Pro Gln Asn Gly
          20             25             30
Leu Tyr Thr Pro Gln Lys Phe Ile Asp Asn Arg Ile Ile Ser Ser Lys
          35             40             45
Tyr Thr Val Trp Asn Phe Val Pro Lys Asn Leu Phe Glu Gln Phe Arg
          50             55             60
Arg Val Ala Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Leu Met
          65             70             75             80
Ile Asp Thr Pro Thr Ser Pro Val Thr Ser Gly Leu Pro Leu Phe Phe
          85             90             95
Val Ile Thr Val Thr Ala Ile Lys Gln Gly Tyr Glu Asp Trp Leu Arg
          100            105            110
His Asn Ser Asp Asn Glu Val Asn Gly Ala Pro Val Tyr Val Val Arg
          115            120            125
Ser Gly Gly Leu Val Lys Thr Arg Ser Lys Asn Ile Arg Val Gly Asp
          130            135            140
Ile Val Arg Ile Ala Lys Asp Glu Ile Phe Pro Ala Asp Leu Val Leu
          145            150            155            160
Leu Ser Ser Asp Arg Leu Asp Gly Ser Cys His Val Thr Thr Ala Ser
          165            170            175
Leu Asp Gly Glu Thr Asn Leu Lys Thr His Val Ala Val Pro Glu Thr

```

180	185	190
Ala Leu Leu Gln Thr Val	Ala Asn Leu Asp Thr Leu Val Ala Val Ile	
195	200	205
Glu Cys Gln Gln Pro Glu	Ala Asp Leu Tyr Arg Phe Met Gly Arg Met	
210	215	220
Ile Ile Thr Gln Gln Met Glu	Glu Ile Val Arg Pro Leu Gly Pro Glu	
225	230	235
Ser Leu Leu Leu Arg Gly	Ala Arg Leu Lys Asn Thr Lys Glu Ile Phe	
245	250	255
Gly Val Ala Val Tyr Thr Gly	Met Glu Thr Lys Met Ala Leu Asn Tyr	
260	265	270
Lys Ser Lys Ser Gln Lys Arg	Ser Ala Val Glu Lys Ser Met Asn Thr	
275	280	285
Phe Leu Ile Ile Tyr Leu Val	Ile Leu Ile Ser Glu Ala Val Ile Ser	
290	295	300
Thr Ile Leu Lys Tyr Thr Trp	Gln Ala Glu Glu Lys Trp Asp Glu Pro	
305	310	315
Trp Tyr Asn Gln Lys Thr Glu	His Gln Arg Asn Ser Ser Lys Ile Leu	
325	330	335
Arg Phe Ile Ser Asp Phe Leu	Ala Phe Leu Val Leu Tyr Asn Phe Ile	
340	345	350
Ile Pro Ile Ser Leu Tyr Val	Thr Val Glu Met Gln Lys Phe Leu Gly	
355	360	365
Ser Phe Phe Ile Gly Trp Asp	Leu Asp Leu Tyr His Glu Glu Ser Asp	
370	375	380
Gln Lys Ala Gln Val Asn Thr	Ser Asp Leu Asn Glu Glu Leu Gly Gln	
385	390	395
Val Glu Tyr Val Phe Thr Asp	Lys Thr Gly Thr Leu Thr Glu Asn Glu	
405	410	415
Met Gln Phe Arg Glu Cys Ser	Ile Asn Gly Met Lys Tyr Gln Glu Ile	
420	425	430
Asn Gly Arg Leu Val Pro Glu	Gly Pro Thr Pro Asp Ser Ser Glu Gly	
435	440	445
Asn Leu Ser Tyr Leu Ser Ser	Leu Ser His Leu Asn Asn Leu Ser His	
450	455	460
Leu Thr Thr Ser Ser Ser Phe	Arg Thr Ser Pro Glu Asn Glu Thr Glu	
465	470	475
Leu Ile Lys Glu His Asp Leu	Phe Phe Lys Ala Val Ser Leu Cys His	
485	490	495
Thr Val Gln Ile Ser Asn Val	Gln Thr Asp Cys Thr Gly Asp Gly Pro	
500	505	510
Trp Gln Ser Asn Leu Ala Pro	Ser Gln Leu Glu Tyr Tyr Ala Ser Ser	
515	520	525
Pro Asp Glu Lys Ala Leu Val	Glu Ala Ala Ala Arg Ile Gly Ile Val	
530	535	540
Phe Ile Gly Asn Ser Glu Glu	Thr Met Glu Val Lys Thr Leu Gly Lys	
545	550	555
Leu Glu Arg Tyr Lys Leu Leu	His Ile Leu Glu Phe Asp Ser Asp Arg	
565	570	575
Arg Arg Met Ser Val Ile Val	Gln Ala Pro Ser Gly Glu Lys Leu Leu	
580	585	590
Phe Ala Lys Gly Ala Glu Ser	Ser Ile Leu Pro Lys Cys Ile Gly Gly	
595	600	605
Glu Ile Glu Lys Thr Arg Ile	His Val Asp Glu Phe Ala Leu Lys Gly	
610	615	620
Leu Arg Thr Leu Cys Ile Ala	Tyr Arg Lys Phe Thr Ser Lys Glu Tyr	

625 630 635 640
 Glu Glu Ile Asp Lys Arg Ile Phe Glu Ala Arg Thr Ala Leu Gln Gln
 645 650 655
 Arg Glu Glu Lys Leu Ala Ala Val Phe Gln Phe Ile Glu Lys Asp Leu
 660 665 670
 Ile Leu Leu Gly Ala Thr Ala Val Glu Asp Arg Leu Gln Asp Lys Val
 675 680 685
 Arg Glu Thr Ile Glu Ala Leu Arg Met Ala Gly Ile Lys Val Trp Val
 690 695 700
 Leu Thr Gly Asp Lys His Glu Thr Ala Val Ser Val Ser Leu Ser Cys
 705 710 715 720
 Gly His Phe His Arg Thr Met Asn Ile Leu Glu Leu Ile Asn Gln Lys
 725 730 735
 Ser Asp Ser Glu Cys Ala Glu Gln Leu Arg Gln Leu Ala Arg Arg Ile
 740 745 750
 Thr Glu Asp His Val Ile Gln His Gly Leu Val Val Asp Gly Thr Ser
 755 760 765
 Leu Ser Leu Ala Leu Arg Glu His Glu Lys Leu Phe Met Glu Val Cys
 770 775 780
 Arg Asn Cys Ser Ala Val Leu Cys Cys Arg Met Ala Pro Leu Gln Lys
 785 790 795 800
 Ala Lys Val Ile Arg Leu Ile Lys Ile Ser Pro Glu Lys Pro Ile Thr
 805 810 815
 Leu Ala Val Gly Asp Gly Ala Asn Asp Val Ser Met Ile Gln Glu Ala
 820 825 830
 His Val Gly Ile Gly Ile Met Gly Lys Glu Gly Arg Gln Ala Ala Arg
 835 840 845
 Asn Ser Asp Tyr Ala Ile Ala Arg Phe Lys Phe Leu Ser Lys Leu Leu
 850 855 860
 Phe Val His Gly His Phe Tyr Tyr Ile Arg Ile Ala Thr Leu Val Gln
 865 870 875 880
 Tyr Phe Phe Tyr Lys Asn Val Cys Phe Ile Thr Pro Gln Phe Leu Tyr
 885 890 895
 Gln Phe Tyr Cys Leu Phe Ser Gln Gln Thr Leu Tyr Asp Ser Val Tyr
 900 905 910
 Leu Thr Leu Tyr Asn Ile Cys Phe Thr Ser Leu Pro Ile Leu Ile Tyr
 915 920 925
 Ser Leu Leu Glu Gln His Val Asp Pro His Val Leu Gln Asn Lys Pro
 930 935 940
 Thr Leu Tyr Arg Asp Ile Ser Lys Asn Arg Leu Leu Ser Ile Lys Thr
 945 950 955 960
 Phe Leu Tyr Trp Thr Ile Leu Gly Phe Ser His Ala Phe Ile Phe Phe
 965 970 975
 Phe Gly Ser Tyr Leu Leu Ile Gly Lys Asp Thr Ser Leu Leu Gly Asn
 980 985 990
 Gly Gln Met Phe Gly Asn Trp Thr Phe Gly Thr Leu Val Phe Thr Val
 995 1000 1005
 Met Val Ile Thr Val Thr Val Lys Met Ala Leu Glu Thr His Phe Trp
 1010 1015 1020
 Thr Trp Ile Asn His Leu Val Thr Trp Gly Ser Ile Ile Phe Tyr Phe
 1025 1030 1035 1040
 Val Phe Ser Leu Phe Tyr Gly Gly Ile Leu Trp Pro Phe Leu Gly Ser
 1045 1050 1055
 Gln Asn Met Tyr Phe Val Phe Ile Gln Leu Leu Ser Ser Gly Ser Ala
 1060 1065 1070
 Trp Phe Ala Ile Ile Leu Met Val Val Thr Cys Leu Phe Leu Asp Ile

1075	1080	1085
Ile Lys Lys Val Phe Asp Arg His Leu His Pro Thr Ser Thr Glu Lys		
1090	1095	1100
Ala Gln Leu Thr Glu Thr Asn Ala Gly Ile Lys Cys Leu Asp Ser Met		
1105	1110	1115
Cys Cys Phe Pro Glu Gly Glu Ala Ala Cys Ala Ser Val Gly Arg Met		
1125	1130	1135
Leu Glu Arg Val Ile Gly Arg Cys Ser Pro Thr His Ile Ser Arg Ser		
1140	1145	1150
Trp Ser Ala Ser Asp Pro Phe Tyr Thr Asn Asp Arg Ser Ile Leu Thr		
1155	1160	1165
Leu Ser Thr Met Asp Ser Ser Thr Cys		
1170	1175	

<210> 3
 <211> 1125
 <212> DNA
 <213> homo sapiens

<400> 3

atgtggcgct	ggatccggca	gcagctgggt	tttgaccac	cacatcagag	tgacacaaga	60
accatctacg	tagccaacag	gtttcctcag	aatggccttt	acacacctca	gaaatttata	120
gataacagga	tcatttcatc	taagtacact	gtgtggaatt	ttgttccaaa	aaattttatt	180
gaacagttca	gaagagtggc	aaacttttat	tttcttatta	tatttttggg	tcagcttatg	240
attgatacac	ctaccagtcc	agttaccagt	ggacttccat	tattctttgt	gataacagta	300
actgccataa	agcagggata	tgaagattgg	ttacggcata	actcagataa	tgaagtaa	360
ggagctcctg	tttatgttgt	tcgaagtggg	ggccttgtaa	aaactagatc	aaaaaacatt	420
cgggtgggtg	atattgttcg	aatagccaaa	gatgaaattt	ttcctgcaga	cttggtgctt	480
ctgtcctcag	atcgactgga	tggttcctgt	cacgttacaa	ctgctagtgt	ggacggagaa	540
actaacctga	agacacatgt	ggcagttcca	gaaacagcat	tattacaaac	agttgccaat	600
ttggacactc	tagtagctgt	aatagaatgc	cagcaaccag	aagcagactt	atacagattc	660
atgggacgaa	tgatcataac	ccaacaaatg	gaagaaattg	taagacctct	ggggccggag	720
agtctcctgc	ttcgtggagc	cagattaaaa	aacacaaaag	aaatttttgg	tggtgcggta	780
tacactggaa	tgaaactaa	gatggcatta	aattacaaga	gcaaatacaca	gaaacgatct	840
gcagtagaaa	agtcaatgaa	tacatttttg	ataatttatc	tagtaattct	tatatctgaa	900
gctgtcatca	gcactatctt	gaagtataca	tggcaagctg	aagaaaaatg	ggatgaacct	960
tggtataacc	aaaaaacaga	acatcaaaga	aatagcaatt	ctgagattta	tttcagactt	1020
ccttgctttt	ttggttctct	acaatttcat	cattccaatt	tcatttatatg	tgacagtcga	1080
aatgcagaaa	tttcttggat	cattttttat	tggctgggat	cttga		1125

<210> 4
 <211> 374
 <212> PRT
 <213> homo sapiens

<400> 4

Met Trp Arg Trp Ile Arg Gln Gln Leu Gly Phe Asp Pro Pro His Gln	
1	15
Ser Asp Thr Arg Thr Ile Tyr Val Ala Asn Arg Phe Pro Gln Asn Gly	
20	30
Leu Tyr Thr Pro Gln Lys Phe Ile Asp Asn Arg Ile Ile Ser Ser Lys	
35	45
Tyr Thr Val Trp Asn Phe Val Pro Lys Asn Leu Phe Glu Gln Phe Arg	
50	60
Arg Val Ala Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Leu Met	
65	80

```
<210> 5
<211> 7277
<212> DNA
<213> homo sapiens
```

<400> 5						
gccgcgggat	gggaacgcgg	cgcggggagt	gaggcagtgg	cggcggcggc	ggtaagcgga	60
acttcggccc	gaggggctcg	cccgcctccg	cctctgtctt	gtcggcctcc	acctgcagcc	120
ccgcggcccc	cgcgccccgc	gggaccgcga	cggcgacgac	gggggaatgt	ggcgctggat	180
ccgcgagcag	ctgggttttg	accaccaca	tcagagtgc	acaagaacca	tctacgtagc	240
caacaggttt	cctcagaatg	gcctttacac	acctcagaaa	tttatagata	acaggatcat	300
ttcatctaag	tacactgtgt	ggaattttgt	tccaaaaaat	ttatttgaa	agttcagaag	360
agtggcaaac	ttttattttc	ttattatatt	tttggttcag	cttatgattg	atacacctac	420
cagtccagtt	accagtggac	ttccattatt	ctttgtgata	acagtaactg	ccataaagca	480
gggatattgaa	gattggttac	ggcataactc	agataatgaa	gtaaatggag	ctcctgttta	540
tgttgttcga	agtggtggcc	ttgtaaaaac	tagatacaaa	aacattcggg	tgggtgatat	600
tggttcgaata	gccaaagatg	aaatttttcc	tgcagacttg	gtgcttctgt	cctcagatcg	660

actggatggt	tcctgtcacg	ttacaactgc	tagtttggac	ggagaaacta	acctgaagac	720
acatgtggca	gttccagaaa	cagcattatt	acaaacagtt	gccaatttgg	acactctagt	780
agctgtaata	gaatgccagc	aaccagaagc	agacttatac	agattcatgg	gacgaatgat	840
cataacccaa	caaattggaag	aaattgtaag	acctctgggg	ccggagagtc	tcctgtctcg	900
tggagccaga	ttaaaaaaca	caaaagaaat	ttttggtgtt	gcggtataca	ctggaatgga	960
aactaagatg	gcattaaatt	acaagagcaa	atcacagaaa	cgatctgcag	tagaaaagtc	1020
aatgaataca	tttttgataa	tttatctagt	aattcttata	tctgaagctg	tcatacagac	1080
tatcttgaag	tatacatggc	aagctgaaga	aaaatgggat	gaaccttggg	ataaccaaaa	1140
aacagaacat	caaagaaata	gcagtaagat	tctgagattt	atttcagact	tccttgcttt	1200
tttggttctc	tacaatttca	tcattccaat	ttcattatat	gtgacagtcg	aaatgcagaa	1260
atttcttgga	tcatttttta	ttggctggga	tcttgatctg	tatcatgaag	aatcagatca	1320
gaaagctcaa	gtcaataactt	ccgatctgaa	tgaagagctt	ggacaggtag	agtacgtgtt	1380
tacagataaa	actggtacac	tgacagaaaa	tgagatgcag	tttcgggaat	gttcaattaa	1440
tggcatgaaa	taccaagaaa	ttaatggtag	acttgtaccc	gaaggaccaa	caccagactc	1500
ttcagaagga	aacttatctt	atcttagtag	tttatcccat	cttaacaaact	tatcccatct	1560
tacaaccagt	tcctctttca	gaaccagtcc	tgaaaatgaa	actgaactaa	ttaaagaaca	1620
tgatctcttc	tttaaagcag	tcagtctctg	tcacactgta	cagattagca	atgttcaaac	1680
tgactgcact	ggtgatggtc	cctggcaatc	caacctggca	ccatcgagtc	tggagtacta	1740
tgcattctca	ccagatgaaa	aggctctagt	agaagctgct	gcaaggattg	gtattgtgtt	1800
tattggcaat	tctgaagaaa	ctatggaggt	taaaactctt	ggaaaactgg	aacggtacaa	1860
actgcttcat	attctggaat	ttgattcaga	tcgtaggaga	atgagtgtaa	ttgttcaggc	1920
accttcagggt	gagaagttaa	tatttgctaa	aggagctgag	tcattcaattc	tcctctaaatg	1980
tatagggtga	gaaatagaaa	aaaccagaat	tcattgtagat	gaatttgctt	tgaaagggct	2040
aagaactctg	tgtatagcat	atagaaaatt	tacatcaaaa	gagtatgagg	aaatagataa	2100
acgcataatt	gaagccagga	ctgccttgca	gcagcgggaa	gagaaattgg	cagctgtttt	2160
ccagttcata	gagaaagacc	tgatattact	tggagccaca	gcagtagaag	acagactaca	2220
agataaagtt	cgagaaacta	ttgaagcatt	gagaatggct	ggtatcaaa	tatgggtact	2280
tactggggat	aaacatgaaa	cagctgttag	tgtgagttta	tcattgtggc	attttcatag	2340
aaccatgaac	atccttgaa	ttataaaacca	gaaatcagac	agcgagtgtg	ctgaacaatt	2400
gaggcagctt	gccagaagaa	ttacagagga	tcattgtgatt	cagcatgggc	tggtagtgga	2460
tgggaccagc	ctatctcttg	cactcagggg	gcattgaaaa	ctattttatg	aagtttgag	2520
aaattgttca	gctgtattat	gctgtcgtat	ggctccactg	cagaaagcaa	aagtaataag	2580
actaataaaa	atatcacctg	agaaacctat	aacattggct	gttggtgatg	gtgctaataga	2640
cgtaagcatg	atacaagaag	cccatgttgg	cataggaatc	atgggtaaag	aaggaagaca	2700
ggctgcaaga	aacagtgcact	atgcaatagc	cagatttaag	ttcctctcca	aattgctttt	2760
tgttcatggt	cattttttatt	atattagaat	agctaccctt	gtacagtatt	ttttttataa	2820
gaatgtgtgc	tttatcacac	cccagttttt	atatcagttc	tactgtttgt	tttctcagca	2880
aacattgtat	gacagcgtgt	acctgacttt	atacaatatt	tgttttactt	ccctacctat	2940
tctgatatat	agtcttttgg	aacagcatgt	agacctcat	gtgttacaaa	ataagcccac	3000
cctttatcga	gacattagta	aaaaccgcct	cttaagtatt	aaaacatttc	tttattggac	3060
catcctgggc	ttcagtcagt	cctttatttt	cttttttggg	tcctattttac	taatagggaa	3120
agatacatct	ctgcttgga	atggccagat	gtygggaaac	tggacatttg	gcactttggt	3180
cttcacagtc	atggttatta	cagtcacagt	aaagatggct	ctggaaactc	atttttggac	3240
ttggatcaac	catctcgtta	cctggggatc	tattatattt	tattttgtat	tttcttgtt	3300
ttatggaggg	attctctggc	catttttggg	ctcccagaat	atgtattttg	tgtttattca	3360
gctcctgtca	agtggttctg	cttggtttgc	cataatcctc	atgggtgtta	catgtctatt	3420
tcttgatata	ataaagaagg	tctttgaccg	acacctccac	cctacaagta	ctgaaaaggc	3480
acagcttact	gaaacaaatg	caggtatcaa	gtgcttggac	tccatgtgct	gtttcccggg	3540
aggagaagca	gcgtgtgcat	ctgttggaag	aatgctggaa	cgagttatag	gaagatgtag	3600
tccaaccac	atcagcagat	catggagtgc	atcggtatcct	ttctatacca	acgacaggag	3660
catcttgact	ctctccacaa	tggaactcat	tacttggtta	aggggcagta	gtactttgtg	3720
ggagccagtt	cacctccttt	cctaaaattc	agtgtgatca	ccctgttaat	ggccacacta	3780
gctctgaaat	taatttccaa	aatctttgta	gtagttcata	cccactcaga	gttataatgg	3840
caaacaata	gaaagcatta	gtacaagccc	ctccaacac	ccttaatttg	aatctgaaca	3900
tgtaaaaatt	tgagaataaa	gagacatttt	tcattctctt	gtctggtttg	tccttgtgtc	3960
ttatgggact	cctaattggca	tttcagtctg	ttgctgaggc	cattatattt	taatataaat	4020

gtagaaaaaa	gagagaaatc	ttagtaaaga	gtatttttta	gtattagctt	gattattgac	4080
tcttctattt	aaatctgctt	ctgtaaatta	tgctgaaagt	ttgccttgag	aactctattt	4140
ttttattaga	gttatattta	aagcttttca	tgggaaaagt	taatgtgaat	actgaggaat	4200
tttggtccct	cagtgacctg	tgttgtaaat	tcattaatgc	attctgagtt	cacagagcaa	4260
attaggagaa	tcatttccaa	ccattattta	ctgcagtatg	gggagtaa	ttataccaat	4320
tcctctaact	gtactgtaac	acagcctgta	aagttagcca	tataaatgca	agggatatatc	4380
atatatacaa	atcaggaatc	aggtccgttc	accgaacttc	aaattgatgt	ttactaatat	4440
ttttgtgaca	gagtataaag	accctatagt	gggtaaaatta	gatactatta	gcatattatt	4500
aaatttaagt	ctttatcatt	ggatcttttg	catgctttta	tctggttaac	atattttaa	4560
ttgctttttt	tctctttacc	tgaaggctct	gtgtatagta	tttcatgaca	tcgttgta	4620
gtttaactat	atcaataaaa	agtttgga	gtattttaa	attgcaaata	tgtttaatta	4680
tacaaatcag	aatagtatgg	gtaattaa	gaatacaaaa	agaagagcct	ctttctgcag	4740
ccgacttaga	catgctcttc	cctttctata	agctagattt	tagaataaag	ggtttcagtt	4800
aataatctta	ttttcaggtt	atgtcatcta	acttatagca	aactaccaca	atcacagtga	4860
ttctgccagt	gtcccagtac	aaggcatatt	tcagggtgtg	ctgtggaatg	taaaaatgct	4920
caacttgat	caggtaatgt	tagcaataaa	ttaaagtcta	agaatgatta	atcgggtaca	4980
tgttactgta	attaactcat	tgcaattcaa	aacctaaact	ccatccctgaa	tttatcaagt	5040
agttcagtat	tgctatttgt	ttttgtttta	ttgaaaagta	atgttgtctt	aagattttaga	5100
agtgattatt	agcttgagaa	ctattacc	gctctaaagca	aataatgatt	gtatacatat	5160
taagataatg	gttaaatgcg	gttttacc	gttttccctt	gaaaatgtaa	ttcctttatg	5220
gagatttatt	gtgcagccct	aagcttccct	cccatttcat	gaatataagg	cttctagaat	5280
tggactggca	ggggaaagaa	tggtagagac	agaaattaag	actttatcct	tgtttgcttg	5340
taaactatta	ttttcttgct	aatgtaacat	ttgtctgttc	cagtgatgta	aggatattaa	5400
gttattaagc	taaatattaa	ttttcaaaaa	tagtccttct	ttaacttaga	tatttcatag	5460
ctggatttag	gaagatctgt	tattctggaa	gtactaaaaa	gaataatata	acgtacaatg	5520
tctgcattca	ctaattcatg	ttccagaaga	ggaaataatg	aagatatact	cagtagagta	5580
ctaggtgga	ggatatggaa	atttgctcat	aaaatctctt	ataaaacgtg	catataacaa	5640
aatgacaccc	agtaggcctg	cattacattt	acatgaccgt	gtttattttg	catcaaataa	5700
actgagtact	gacaccagac	aaagactcca	aagtcataaa	atagcctatg	accaactgca	5760
gcaagacagg	aggtcagctc	gcctataatg	gtgcttaaa	tggtgattgat	gtaattttct	5820
gtactacca	tttgaagtta	gttaaggaga	actttatttt	tttaaaaaaa	gtaaatggca	5880
accactagt	tgctcatcct	gaactgttac	tccaaatcca	ctccgttttt	aaagcaaaat	5940
tatcttgtga	ttttaagaaa	agagttttct	atttatttaa	gaaagtaaca	atgcagtctg	6000
caagctttca	gtagttttct	agtgtatat	tcactctgta	aaactcttac	tacgtaacca	6060
gtaatcacia	ggaaagtgtc	ccctttgcat	atttctttta	aattctttct	ttggaaagta	6120
tgatgttgat	aattaactta	cccttatctg	ccaaaaccag	agcaaaatgc	taaatacgtt	6180
attgctaata	agtggtctca	aatcgatttg	cctccctttg	cctcgtctga	gggctgtaag	6240
cctgaagata	gtggcaagca	caaagtcagt	ttccaaaatt	gccccctcag	tgctttaagt	6300
gactcagcac	cctgcctcag	cttcagcagg	cstaggctca	ccctgggcgg	agcaaagtat	6360
gggccaggga	gaactacagc	tacgaagacc	tgctgtcgag	ttgagaaaag	gggagaattt	6420
atggtctgaa	ttttctaact	gtcctctttc	ttgggtctaa	agctcataat	acacaaaggc	6480
ttccagacct	gagccacacc	caggccctat	cctgaacagg	agactaaaca	gaggcaaatc	6540
aaccctagga	aataacttgca	ttctgcccta	cggttagtac	caggactgag	gtcatttcta	6600
ctggaaaaga	ttgtgagatt	gaacttatct	gatcgcttga	gactccta	aggcaggagt	6660
caaggccact	agaaaattga	cagttaagag	ccaaaagt	ttaaaatatg	ctactctgaa	6720
aaatctcgtg	aaggctgtag	gaaaaggag	aatcttccat	gttggtgttt	ttcctgtaaa	6780
gatcagtttg	gggtatgata	taagcaggta	ttaataaaaa	taacacacca	aagagttacg	6840
taaaacatgt	tttattaatt	ttggtcccca	cgtacagaca	ttttatttct	attttgaaat	6900
gagttatcta	ttttcataaa	agtaaaacac	tattaaagt	ctgttttatg	tgaaataact	6960
tgaatgttgt	tcctataaaa	aatagatcat	aactcatgat	atgtttgtaa	tcatggta	7020
ttagattttt	atgaggaatg	agtatctgga	aatattgtag	caatacttgg	tttaaaattt	7080
tggaacctgag	acactgtggc	tgtctaagt	aatcctttta	aaattctctg	cattgtcagt	7140
aaatgtagta	tattattgta	cagctactca	taatttttta	aagtttatga	agttatattt	7200
atcaataaaa	aactttccta	tataattaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaacaaa	7260
aaaaaaaaaa	aaaaaaa					7277

<210> 6
 <211> 2913
 <212> DNA
 <213> homo sapiens

<400> 6
 atggcacaac tagagaggag cgccatctct ggcttcagct ctaagtccag gcgaaactca 60
 ttgcataatg atgttaagcg tgaagtatac aatgaggaga cctttcaaca ggaacacaaa 120
 aggaaggcct cctcttctgg gaacatgaac atcaacatca ccaccttcag acaccacgtc 180
 cagtgccgct gctcatggca caggttccta cgatgcertgc ttacaatctt tcccttccta 240
 gaatggatgt gtatgtatcg attaaaggat tggcttctgg gagacttact tgctggtata 300
 agtgttggcc ttgtgcaagt tccccaaggc ctgacactta gtttgctggc aaggcaactg 360
 attcctcctc tcaacatcgc ttatgcagct ttctgttctt cggtaatcta tgtaattttt 420
 ggatcgtgtc atcaaatgtc cgttggttcc ttcttcctgg tgagtgtctt gctgatcaac 480
 gttctgaaag tgagcccatt caacaacggt caactgggtca tgggatcttt cgtcaagaat 540
 gagttttcgg cccctccta ccttatgggc tataataaat ccttgagtgt ggtggcaacc 600
 acaacttttc tgactgggat tattcagcta ataatgggcg tattgggttt gggcttcatt 660
 gccacttacc ttccggagtc tgcaatgaat gcttacctgg ctgctgtggc acttcataac 720
 atgctgtccc agctgacttt catccttggg attatgatta gtttccatgc cgggtccatc 780
 tccttcttct atgacataat taattactgt gtatgtctcc caaagcgaa ttccaccagc 840
 attctagtat ttctaactgt tgttgttgtc ctgcgaatca acaaatgtat cagaatttct 900
 ttcaatcagt atcccattga gtttcccatg gaattatttc tgattattgg cttcactgtg 960
 attgcaaca agataagcat ggccacagaa accagccaga cgcttattga catgattcct 1020
 tatagctttc tgcttctgt aacaccagat ttcagccttc ttcccaagat aattttacaa 1080
 gccttctcct tatctttggt gagctccttt ctgctcatat ttctgggcaa gaagattgcc 1140
 agtcttcaca attacagtgt caattccaac caggatttaa tagccatcgg cctttgcaat 1200
 gtcgtcagtt catttttcag atcttgtgtg tttactgggt ctattgctag gactattatc 1260
 caggataaat ctggaggag acaacagttt gcatctctgg taggcgcagg tgtgatgctg 1320
 ctcctgatgg tgaagatggg acactttttc tacacactgc caaatgctgt gctggctggt 1380
 attattctga gcaacgtcat tccctacctt gaaaccattt ctaacctacc cagcctgtgg 1440
 aggcaggacc aatatgactg tgctctttgg atgatgacat tctcatcttc aattttcctg 1500
 ggactggaca ttggactaat tatctcagta gtttctgctt tcttcatcac cactgttcgt 1560
 tcacacagag ctaagattct tctcctgggt caaatcccta acaccaacat ttatagaagc 1620
 atcaatgatt atcgggagat catcaccatt cctgggggtga aaatcttcca gtgctgcagc 1680
 tcaattacat ttgtaaagt ttactaccta aagcataagc tgtaaaaaga ggtgatatg 1740
 gtaaagggtg ctcttaaaga agaagaaatt ttcagcttgt ttaattcaag tgacaccaat 1800
 ctacaaggag gaaagatttg caggtgtttc tgcaactgtg atgatctgga gccgctgccc 1860
 aggattcttt acacagagcg atttgaaaat aaactggatc ccgaagcatc ctccattaac 1920
 ctgattcact gctcacattt tgagagcatg aacacaagcc aaactgcatc cgaagaccaa 1980
 gtgccatata cagtatcgtc cgtgtctcag aaaaatcaag ggcaacagta tgaggagggtg 2040
 gaggaagtgt ggcttcctaa taactcatca agaaacagct caccaggact gcctgatgtg 2100
 gcggaagccc aggggaggag atcactcatc ccttactcag atgctgtctt actgcccagt 2160
 gtccacacca tcatcctgga tttctccatg gtacactacg tggattcacg ggggttagtc 2220
 gtattaagac agatatgcaa tgcctttcaa aacgccaaca ttttgatact cattgcaggg 2280
 tgtcactctt ccatagtcag ggcatttgag aggaatgatt tctttgacgc tggcatcacc 2340
 aagaccagc tgctcctcag cgttcacgac gccgtgctgt ttgccttgtc aaggaagggtc 2400
 ataggctcct ctgagttaag catcgatgaa tccgagacag tgatacggga aacctactca 2460
 gaaacagaca agaattgaaa ttcaagatat aaaatgagca gcagttttct aggaagccaa 2520
 aaaaatgtaa gtccaggctt catcaagatc caacagcctg tagaagagga gtcggagtgtg 2580
 gatttgagc tggaatcaga acaagaggct gggctgggtc tggacctaga cctggatcgg 2640
 gagctggagc ctgaaatgga gcccaaggct gagaccaga ccaagacca gaccgagatg 2700
 gagccccagc ctgagactga gcctgagatg gagcccaacc ccaaacttag gccaaagagct 2760
 cacacttttc ctcagcagcg ttactggcct atgtatcatc cgtctatggc ttccaccagc 2820
 tctcagactc agactcggac atggtcagtg gagaggagac gccatcctat ggattcatac 2880
 tcaccagagg gcaacagcaa tgaagatgtc tag 2913

<210> 7
 <211> 970
 <212> PRT
 <213> homo sapiens

<220>
 <221> VARIANT
 <222> (1)...(970)
 <223> Xaa = Any Amino Acid

<400> 7
 Met Ala Gln Leu Glu Arg Ser Ala Ile Ser Gly Phe Ser Ser Lys Ser
 1 5 10 15
 Arg Arg Asn Ser Phe Ala Tyr Asp Val Lys Arg Glu Val Tyr Asn Glu
 20 25 30
 Glu Thr Phe Gln Gln Glu His Lys Arg Lys Ala Ser Ser Ser Gly Asn
 35 40 45
 Met Asn Ile Asn Ile Thr Thr Phe Arg His His Val Gln Cys Arg Cys
 50 55 60
 Ser Trp His Arg Phe Leu Arg Cys Met Leu Thr Ile Phe Pro Phe Leu
 65 70 75 80
 Glu Trp Met Cys Met Tyr Arg Leu Lys Asp Trp Leu Leu Gly Asp Leu
 85 90 95
 Leu Ala Gly Ile Ser Val Gly Leu Val Gln Val Pro Gln Gly Leu Thr
 100 105 110
 Leu Ser Leu Leu Ala Arg Gln Leu Ile Pro Pro Leu Asn Ile Ala Tyr
 115 120 125
 Ala Ala Phe Cys Ser Ser Val Ile Tyr Val Ile Phe Gly Ser Cys His
 130 135 140
 Gln Met Ser Val Gly Ser Phe Phe Leu Val Ser Ala Leu Leu Ile Asn
 145 150 155 160
 Val Leu Lys Val Ser Pro Phe Asn Asn Gly Gln Leu Val Met Gly Ser
 165 170 175
 Phe Val Lys Asn Glu Phe Ser Ala Pro Ser Tyr Leu Met Gly Tyr Asn
 180 185 190
 Lys Ser Leu Ser Val Val Ala Thr Thr Thr Phe Leu Thr Gly Ile Ile
 195 200 205
 Gln Leu Ile Met Gly Val Leu Gly Leu Gly Phe Ile Ala Thr Tyr Leu
 210 215 220
 Pro Glu Ser Ala Met Asn Ala Tyr Leu Ala Ala Val Ala Leu His Ile
 225 230 235 240
 Met Leu Ser Gln Leu Thr Phe Ile Phe Gly Ile Met Ile Ser Phe His
 245 250 255
 Ala Gly Pro Ile Ser Phe Phe Tyr Asp Ile Ile Asn Tyr Cys Val Ala
 260 265 270
 Leu Pro Lys Ala Asn Ser Thr Ser Ile Leu Val Phe Leu Thr Val Val
 275 280 285
 Val Ala Leu Arg Ile Asn Lys Cys Ile Arg Ile Ser Phe Asn Gln Tyr
 290 295 300
 Pro Ile Glu Phe Pro Met Glu Leu Phe Leu Ile Ile Gly Phe Thr Val
 305 310 315 320
 Ile Ala Asn Lys Ile Ser Met Ala Thr Glu Thr Ser Gln Thr Leu Ile
 325 330 335
 Asp Met Ile Pro Tyr Ser Phe Leu Leu Pro Val Thr Pro Asp Phe Ser
 340 345 350
 Leu Leu Pro Lys Ile Ile Leu Gln Ala Phe Ser Leu Ser Leu Val Ser

355	360	365
Ser Phe Leu Leu Ile Phe Leu Gly Lys Lys Ile Ala Ser Leu His Asn		
370	375	380
Tyr Ser Val Asn Ser Asn Gln Asp Leu Ile Ala Ile Gly Leu Cys Asn		
385	390	395
Val Val Ser Ser Phe Phe Arg Ser Cys Val Phe Thr Gly Ala Ile Ala		400
	405	410
Arg Thr Ile Ile Gln Asp Lys Ser Gly Gly Arg Gln Gln Phe Ala Ser		415
	420	425
Leu Val Gly Ala Gly Val Met Leu Leu Leu Met Val Lys Met Gly His		430
	435	440
Phe Phe Tyr Thr Leu Pro Asn Ala Val Leu Ala Gly Ile Ile Leu Ser		445
	450	455
Asn Val Ile Pro Tyr Leu Glu Thr Ile Ser Asn Leu Pro Ser Leu Trp		460
465	470	475
Arg Gln Asp Gln Tyr Asp Cys Ala Leu Trp Met Met Thr Phe Ser Ser		480
	485	490
Ser Ile Phe Leu Gly Leu Asp Ile Gly Leu Ile Ile Ser Val Val Ser		495
	500	505
Ala Phe Phe Ile Thr Thr Val Arg Ser His Arg Ala Lys Ile Leu Leu		510
	515	520
Leu Gly Gln Ile Pro Asn Thr Asn Ile Tyr Arg Ser Ile Asn Asp Tyr		525
	530	535
Arg Glu Ile Ile Thr Ile Pro Gly Val Lys Ile Phe Gln Cys Cys Ser		540
545	550	555
Ser Ile Thr Phe Val Asn Val Tyr Tyr Leu Lys His Lys Leu Leu Lys		560
	565	570
Glu Val Asp Met Val Lys Val Pro Leu Lys Glu Glu Glu Ile Phe Ser		575
	580	585
Leu Phe Asn Ser Ser Asp Thr Asn Leu Gln Gly Gly Lys Ile Cys Arg		590
	595	600
Cys Phe Cys Asn Cys Asp Asp Leu Glu Pro Leu Pro Arg Ile Leu Tyr		605
	610	615
Thr Glu Arg Phe Glu Asn Lys Leu Asp Pro Glu Ala Ser Ser Ile Asn		620
625	630	635
Leu Ile His Cys Ser His Phe Glu Ser Met Asn Thr Ser Gln Thr Ala		640
	645	650
Ser Glu Asp Gln Val Pro Tyr Thr Val Ser Ser Val Ser Gln Lys Asn		655
	660	665
Gln Gly Gln Gln Tyr Glu Glu Val Glu Glu Val Trp Leu Pro Asn Asn		670
	675	680
Ser Ser Arg Asn Ser Ser Pro Gly Leu Pro Asp Val Ala Glu Ser Gln		685
	690	695
Gly Arg Arg Ser Leu Ile Pro Tyr Ser Asp Ala Ser Leu Leu Pro Ser		700
705	710	715
Val His Thr Ile Ile Leu Asp Phe Ser Met Val His Tyr Val Asp Ser		720
	725	730
Arg Gly Leu Val Val Leu Arg Gln Ile Cys Asn Ala Phe Gln Asn Ala		735
	740	745
Asn Ile Leu Ile Leu Ile Ala Gly Cys His Ser Ser Ile Val Arg Ala		750
	755	760
Phe Glu Arg Asn Asp Phe Phe Asp Ala Gly Ile Thr Lys Thr Gln Leu		765
	770	775
Phe Leu Ser Val His Asp Ala Val Leu Phe Ala Leu Ser Arg Lys Val		780
785	790	795
Ile Gly Ser Ser Glu Leu Ser Ile Asp Glu Ser Glu Thr Val Ile Arg		800

805										810					815					
Glu	Thr	Tyr	Ser	Glu	Thr	Asp	Lys	Asn	Asp	Asn	Ser	Arg	Tyr	Lys	Met					
820										825					830					
Ser	Ser	Ser	Phe	Leu	Gly	Ser	Gln	Lys	Asn	Val	Ser	Pro	Gly	Phe	Ile					
835										840					845					
Lys	Ile	Gln	Gln	Pro	Val	Glu	Glu	Glu	Ser	Glu	Leu	Asp	Leu	Glu	Leu					
850										855					860					
Glu	Ser	Glu	Gln	Glu	Ala	Gly	Leu	Gly	Leu	Asp	Leu	Asp	Leu	Asp	Arg					
865										870					875					
Glu	Leu	Glu	Pro	Glu	Met	Glu	Pro	Lys	Ala	Glu	Thr	Glu	Thr	Lys	Thr					
885										890					895					
Gln	Thr	Glu	Met	Glu	Pro	Gln	Pro	Glu	Thr	Glu	Pro	Glu	Met	Glu	Pro					
900										905					910					
Asn	Pro	Lys	Ser	Arg	Pro	Arg	Ala	His	Thr	Phe	Pro	Gln	Gln	Arg	Tyr					
915										920					925					
Trp	Pro	Met	Tyr	His	Pro	Ser	Met	Ala	Ser	Thr	Gln	Ser	Gln	Thr	Gln					
930										935					940					
Thr	Arg	Thr	Trp	Ser	Val	Glu	Arg	Arg	Arg	His	Pro	Met	Asp	Ser	Tyr					
945										950					955					
Ser	Pro	Glu	Gly	Asn	Ser	Asn	Glu	Asp	Val											
965										970										

<210> 8

<211> 3749

<212> DNA

<213> homo sapiens

<400> 8

ttttccaact	ccccatctcc	tcctctctca	gattaaaaga	agttatatgg	actttgtgat	60
gttttctgcc	gctttgtgaa	gtaggcctta	tttctcttgt	cctttcgtac	agggaggaaat	120
ttgaagtaga	tagaaaccga	cctggattac	tccggtctga	actcagatca	cgtaggactt	180
taatcgttga	acaaacgaac	ctttaatagc	ggctgcacca	tcgggatgtc	ctgatccaac	240
atcgaggtcg	taaaccctat	tggtgatatg	gactctagaa	taggattgcg	ctgttatccc	300
tagggtaact	tgttccgttg	gtcaagttat	tggatcaatt	gagtatagta	gttcgctttg	360
actggtgaag	tcttgcatg	tactgctcgg	aggttgggtt	ctgctccgag	gtcgccccaa	420
ccgaaatttt	taatgcagga	gcgcccgcac	tcccgcctcc	gccaaggagc	caggaatggc	480
acaactagag	aggagcgcca	tctctggctt	cagctctaag	tccaggcgaa	actcattcgc	540
atatgatgtt	aagcgtgaag	tatacaatga	ggagaccttt	caacaggaac	acaaaaggaa	600
ggcctcctct	tctgggaaca	tgaacatcaa	catcaccacc	ttcagacacc	acgtccagt	660
ccgctgctca	tggcacaggt	tcctacgatg	crtgcttaca	atctttccct	tcctagaatg	720
gagtggtatg	tatcgattaa	aggattggct	tctgggagac	ttacttgctg	gtataagtgt	780
tggccttggtg	caagtccccc	aaggcctgac	acttagtttg	ctggcaaggc	aactgattcc	840
tcctctcaac	atcgcttatg	cagctttctg	ttcttcggta	atctatgtaa	tttttggtatc	900
gtgtcatcaa	atgtccgttg	gttccttctt	cctgggtgagt	gctctgctga	tcaacgttct	960
gaaagtgagc	ccattcaaca	acggccaact	ggcatggga	tctttcgtca	agaatgagtt	1020
ttcggccccc	tcctacctta	tgggctataa	taaacccttg	agtggtgggtg	caaccacaac	1080
ttttctgact	gggattattc	agctaataat	gggcgtattg	ggtttgggtg	tcattgccac	1140
ttaccttcg	gagtcctgaa	tgaatgctta	cctggctgct	gtggcacttc	atatcatgct	1200
gtcccagctg	actttcatct	ttgggattat	gattagtttc	catgccggtc	ccatctcctt	1260
cttctatgac	ataattaatt	actgtgtagc	tctcccaaaa	gcgaattcca	ccagcattct	1320
agtattttcta	actgttgttg	ttgctctgcg	aatcaacaaa	tgtatcagaa	tttctttcaa	1380
tcagtatccc	attgagtttc	ccatggaatt	atttctgatt	attggcttca	ctgtgattgc	1440
aaacaagata	agcatggcca	cagaaaccag	ccagacgctt	attgacatga	ttccttatag	1500
ctttctgctt	cctgtaaacac	cagatttcag	ccttcttccc	aagataattt	tacaagcctt	1560
ctccttatct	ttgggtgagct	cctttctgct	catatttctg	ggcaagaaga	ttgccagtct	1620
tcacaattac	agtgtcaatt	ccaaccagga	tttaatagcc	atcggccttt	gcaatgtcgt	1680

cagttcattt	ttcagatcct	gtgtgtttac	tggtgctatt	gctaggacta	ttatccagga	1740
taaactctga	ggaagacaac	agtttgcac	tctggtaggc	gcaggtgtga	tgctgctcct	1800
gatggtgaag	atgggacact	ttttctacac	actgccaaat	gctgtgctgg	ctgggtattat	1860
tctgagcaac	gtcattccct	accttgaaac	catttctaac	ctacccagcc	tgaggaggca	1920
ggaccaatat	gactgtgctc	tttgatgat	gacattctca	tcttcaattt	tcctgggact	1980
ggacattgga	ctaattatct	cagtagtttc	tgctttcttc	atcaccactg	ttcgttcaca	2040
cagagctaag	attcttctcc	tgggtcaaat	ccctaacacc	aacatttata	gaagcatcaa	2100
tgattatcgg	gagatcatca	ccattcctgg	ggtgaaaatc	ttccagtgc	gcagctcaat	2160
tacatttgta	aatgtttact	acctaagca	taagctgtta	aaagagggtg	atatggtaaa	2220
ggtgcctctt	aaagaagaag	aaattttcag	cttgtttaat	tcaagtgaca	ccaatctaca	2280
aggaggaaag	atttgcaagg	gtttctgcaa	ctgtgatgat	ctggagccgc	tgcccaggat	2340
tctttacaca	gagcgatttg	aaaataaact	ggatcccga	gcatcctcca	ttaacctgat	2400
tactgtctca	cattttgaga	gcatgaacac	aagccaaact	gcatccgaag	accaagtgcc	2460
atacacagta	tcgtccgtgt	ctcagaaaaa	tcaagggcaa	cagtatgagg	aggtggagga	2520
agtttggtct	cctaataact	catcaagaaa	cagctcacca	ggactgcctg	atgtggcgga	2580
aagccagggg	aggagatcac	tcacccctta	ctcagatgcg	tctctactgc	ccagtgtcca	2640
caccatcatc	ctggatttct	ccatggtaca	ctacgtggat	tcacgggggt	tagtcgtatt	2700
aagacagata	tgcaatgcct	ttcaaaacgc	caacattttg	atactcattg	caggggtgca	2760
ctctccata	gtcagggcat	ttgagaggaa	tgatttcttt	gacgtggca	tcaccaagac	2820
ccagctgttc	ctcagcgttc	acgacgccgt	gctgtttgcc	ttgtcaagga	aggcatagag	2880
ctcctctgag	ttaagcatcg	atgaatccga	gacagtata	cgggaaacct	actcagaaac	2940
agacaagaat	gacaattcaa	gatataaaat	gagcagcagt	tttctaggaa	gccaaaaaaa	3000
tgtaagtcca	ggcttcatca	agatccaaca	gcctgtagaa	gaggagtcgg	agttggattt	3060
ggagctggaa	tcagaacaag	aggctgggct	gggtctggac	ctagacctgg	atcgggagct	3120
ggagcctgaa	atggagccca	aggctgagac	cgagaccaag	acccagaccg	agatggagcc	3180
ccagcctgag	actgagcctg	agatggagcc	caaccccaaa	tctaggccaa	gagctcacac	3240
ttttcctcag	cagcgttact	ggcctatgta	tcacccgtct	atggcttcca	cccagtctca	3300
gactcagact	cgacatggt	cagtggagag	gagacgccat	cctatggatt	catactcacc	3360
agagggcaac	agcaatgaag	atgtctagga	gatgaactag	aaataagggg	tcagataatg	3420
ctggcaaatc	ctcctaccca	aaaaggggtc	aattgtccag	agacctagac	tggtacgaa	3480
ctagcagtac	ttccttctctg	actgtgactc	ctactacctg	ccagccttct	tccttctct	3540
gcgtgggat	catactccca	aatcacatta	ctaaatgccca	acaattatct	ctgaattccc	3600
tatccaggct	cccctcattt	caccttcagc	atatattcta	gtcatgaatt	tccttcttca	3660
cacacccac	atctctgggc	tttgtgccag	accatctcta	acttaatcct	ctcatccctg	3720
ttccccttct	tccaaagaga	tgaagctca				3749

<210> 9

<211> 1524

<212> DNA

<213> homo sapiens

<400> 9

atgggggtgt	ggggtcggaa	ccggggcccg	ctgctgtgca	tgctggcgct	gaccttcatg	60
ttcatggtgc	tgaggtggt	ggtgagccgg	gtgacctcgt	cgctggcgat	gctctccgac	120
tccttccaca	tgctgtcgga	cgtgctggcg	ctggtggtgg	cgctggtggc	cgagcgcttc	180
gcccggcgga	cccacgccac	ccagaagaac	acgttcggct	ggatccgagc	cgaggtaatg	240
ggggctcttg	tgaacgccat	cttctgact	ggcctctgtt	tcgccatcct	gctggaggcc	300
atcgagcgct	tcacgagcc	gcacgagatg	cagcagccgc	tggtggtcct	tggggctcggc	360
gtggccgggc	tgctggtcaa	cgtgctgggg	ctctgcctct	tccaccatca	cagcggtctc	420
agccaggact	ccggccacgg	ccactcgcac	gggggtcacg	gccacggcca	cgccctcccc	480
aaggggcctc	gcgttaagag	cacccgcccc	gggagcagcg	acatcaacgt	ggccccgggc	540
gagcagggtc	ccgaccagga	ggagaccaac	accctgggtg	ccaataccag	caactccaac	600
gggctgaaat	tggaacccgc	agaccagaa	aacccagaa	gtggtgatac	agtggaaagta	660
caagtgaatg	gaaatcttgt	cagagaacct	gacccatagg	aactggaaga	agatagggtc	720
ggacaactta	catgcgtgg	agtttttctg	catgtccttg	gagatgcctt	gggttcagtg	780
attgtagtag	taaatgcctt	agtcttttac	ttttcttggg	aaggttggtc	tgaaggggat	840

```

ttttgtgtga atccatgttt ccctgacccc tgcaaagcat ttgtagaaat aattaatagt 900
actcatgcat cactttatga ggctgggtcct tgctgggtgc tatatttaga tccaactctt 960
tgtgttgtaa tgggtttgat acttctttac acaacctatc cattacttaa ggaatctgct 1020
cttattcttc tacaaactgt tcctaaacaa attgatatca gaaatttgat aaaagaactt 1080
cgaaatgttg aaggagttag ggaagttcat gaattacatg tttggcaact tgctggaagc 1140
agaatcattg ccactgctca cataaaatgt gaagatccaa catcatacat ggaggtggct 1200
aaaaccatta aagacgtttt tcataatcac ggaattcacg ctactaccat tcagcctgaa 1260
tttgctagtg taggctctaa atcaagtgtg gttccgtgtg aacttgcctg cagaacccag 1320
tgtgctttga agcaatgttg tgggacacta ccacaagccc cttatggaaa ggatgcagaa 1380
aagaccccag cagttagcat ttcttgttta gaacttagta acaatctaga gaagaagccc 1440
aggaggacta aagctgaaaa catccctgct gttgtgatag agattaaaaa catgccaac 1500
aaacaacctg aatcatcttt gtga 1524

```

<210> 10

<211> 507

<212> PRT

<213> homo sapiens

<400> 10

```

Met Gly Cys Trp Gly Arg Asn Arg Gly Arg Leu Leu Cys Met Leu Ala
1          5          10          15
Leu Thr Phe Met Phe Met Val Leu Glu Val Val Val Ser Arg Val Thr
20          25          30
Ser Ser Leu Ala Met Leu Ser Asp Ser Phe His Met Leu Ser Asp Val
35          40          45
Leu Ala Leu Val Val Ala Leu Val Ala Glu Arg Phe Ala Arg Arg Thr
50          55          60
His Ala Thr Gln Lys Asn Thr Phe Gly Trp Ile Arg Ala Glu Val Met
65          70          75          80
Gly Ala Leu Val Asn Ala Ile Phe Leu Thr Gly Leu Cys Phe Ala Ile
85          90          95
Leu Leu Glu Ala Ile Glu Arg Phe Ile Glu Pro His Glu Met Gln Gln
100         105         110
Pro Leu Val Val Leu Gly Val Gly Val Ala Gly Leu Leu Val Asn Val
115         120         125
Leu Gly Leu Cys Leu Phe His His His Ser Gly Phe Ser Gln Asp Ser
130         135         140
Gly His Gly His Ser His Gly Gly His Gly His Gly His Gly Leu Pro
145         150         155         160
Lys Gly Pro Arg Val Lys Ser Thr Arg Pro Gly Ser Ser Asp Ile Asn
165         170         175
Val Ala Pro Gly Glu Gln Gly Pro Asp Gln Glu Glu Thr Asn Thr Leu
180         185         190
Val Ala Asn Thr Ser Asn Ser Asn Gly Leu Lys Leu Asp Pro Ala Asp
195         200         205
Pro Glu Asn Pro Arg Ser Gly Asp Thr Val Glu Val Gln Val Asn Gly
210         215         220
Asn Leu Val Arg Glu Pro Asp His Met Glu Leu Glu Glu Asp Arg Ala
225         230         235         240
Gly Gln Leu Asn Met Arg Gly Val Phe Leu His Val Leu Gly Asp Ala
245         250         255
Leu Gly Ser Val Ile Val Val Val Asn Ala Leu Val Phe Tyr Phe Ser
260         265         270
Trp Lys Gly Cys Ser Glu Gly Asp Phe Cys Val Asn Pro Cys Phe Pro
275         280         285
Asp Pro Cys Lys Ala Phe Val Glu Ile Ile Asn Ser Thr His Ala Ser

```

290 295 300
 Leu Tyr Glu Ala Gly Pro Cys Trp Val Leu Tyr Leu Asp Pro Thr Leu
 305 310 315 320
 Cys Val Val Met Val Cys Ile Leu Leu Tyr Thr Thr Tyr Pro Leu Leu
 325 330 335
 Lys Glu Ser Ala Leu Ile Leu Leu Gln Thr Val Pro Lys Gln Ile Asp
 340 345 350
 Ile Arg Asn Leu Ile Lys Glu Leu Arg Asn Val Glu Gly Val Glu Glu
 355 360 365
 Val His Glu Leu His Val Trp Gln Leu Ala Gly Ser Arg Ile Ile Ala
 370 375 380
 Thr Ala His Ile Lys Cys Glu Asp Pro Thr Ser Tyr Met Glu Val Ala
 385 390 395 400
 Lys Thr Ile Lys Asp Val Phe His Asn His Gly Ile His Ala Thr Thr
 405 410 415
 Ile Gln Pro Glu Phe Ala Ser Val Gly Ser Lys Ser Ser Val Val Pro
 420 425 430
 Cys Glu Leu Ala Cys Arg Thr Gln Cys Ala Leu Lys Gln Cys Cys Gly
 435 440 445
 Thr Leu Pro Gln Ala Pro Tyr Gly Lys Asp Ala Glu Lys Thr Pro Ala
 450 455 460
 Val Ser Ile Ser Cys Leu Glu Leu Ser Asn Asn Leu Glu Lys Lys Pro
 465 470 475 480
 Arg Arg Thr Lys Ala Glu Asn Ile Pro Ala Val Val Ile Glu Ile Lys
 485 490 495
 Asn Met Pro Asn Lys Gln Pro Glu Ser Ser Leu
 500 505

<210> 11

<211> 2222

<212> DNA

<213> homo sapiens

<400> 11

ctccggctgc ggctcttggt accccggctc cgggagccca gctccccgcc accgccgccg 60
 cctgggtgtg ggggctgctg aggctgagcc gggcttcggc gccggctctg aggacggacg 120
 cctgaggagc tgcgcggcgc ggcgccgccg gctggcggag aacgcccaca ggcgcggggc 180
 tcggcggtt gaccgggct tgtccccgtg cggccgcggg ggcccctcag cggtttccccg 240
 aacggcccga ctccggcgct cctccgtgtc gcggtgcgcg accctccgcg tcccgccaac 300
 gccgccgctg caccagtctc cgggccgggc tcggcggggc ccgcagccgc agccatgggg 360
 tgttggggtc ggaaccgggg ccggctgctg tgcagtctgg cgctgacctt catgttcatg 420
 gtgctggagg tgggtgtgag ccgggtgacc tcgtcgctgg cgatgctctc cgactccttc 480
 cacatgtgt cggacgtgct ggcgtgtgtg gtggcgctgg tggccgagcg cttcgccccg 540
 cggaccacg ccaccagaa gaacacgttc ggctggatcc gagccgaggt aatgggggct 600
 ctggtgaacg ccattcttct gactggcctc tgtttcgcca tcctgctgga ggccatcgag 660
 cgcttcatcg agccgcacga gatgcagcag ccgctggtgg tccttggggg cggcgtggcc 720
 gggctgctgg tcaacgtgct ggggctctgc ctctccacc atcacagcgg cttcagccag 780
 gactccggcc acggccactc gcacgggggt cacggccacg gccacggcct cccaagggg 840
 cctcgctta agagcaccg ccccgggagc agcgacatca acgtggcccc gggcgagcag 900
 ggtcccgacc aggaggagac caacaccctg gtggccaata ccagcaactc caacgggctg 960
 aaattggacc ccgcagaccc agaaaacccc agaagtgtg atacagtgga agtacaagtg 1020
 aatggaaatc ttgtcagaga acctgaccat atggaactgg aagaagatag ggctggacaa 1080
 cttaacatgc gtggagtttt tctgcatgtc cttggagatg ccttgggttc agtgattgta 1140
 gtagtaaatg ccttagtctt ttacttttct tggaagggtt gttctgaagg ggatttttgt 1200
 gtgaatccat gtttccctga cccctgcaaa gcattttag aaataattaa tagtactcat 1260
 gcatcacttt atgaggctgg tccttgctgg gtgctatatt tagatccaac tctttgtgtt 1320

gtaatggttt	gtatacttct	ttacacaacc	tatccattac	ttaaggaatc	tgctcttatt	1380
cttctacaaa	ctgttcctaa	acaaattgat	atcagaaatt	tgataaaaaga	acttcgaaat	1440
ggtgaaggag	ttgaggaagt	tcatgaatta	catgtttggc	aacttgctgg	aagcagaatc	1500
attgccactg	ctcacataaa	atgtgaagat	ccaacatcat	acatggaggt	ggctaaaacc	1560
attaaagacg	tttttcataa	tcacggaatt	cacgctacta	ccattcagcc	tgaatttgct	1620
agtgtaggct	ctaaatcaag	tgtagtccg	tgtgaacttg	cctgcagaac	ccagtgtgct	1680
ttgaagcaat	gttgtgggac	actaccacaa	gccccttatg	gaaaggatgc	agaaaagacc	1740
ccagcagtta	gcatttcttg	tttagaactt	agtaacaatc	tagagaagaa	gcccaggagg	1800
actaaagctg	aaaacatccc	tgctgttggtg	atagagatta	aaaacatgcc	aaacaaacaa	1860
cctgaatcat	ctttgtgagt	cttgaaaaag	atgtgatatt	tgacttttgc	tttaaactgc	1920
aagaggaaaa	agactccact	gaaattctaa	gtttgccaag	tagtgtaatt	gaagtccttg	1980
tctggtcaca	cagtttaatt	ctatttttgt	aagaacataa	tgggactgca	taacagagtt	2040
ctatattaca	atgtgtgatt	attagtacag	agtacagcta	tgctgtgact	gttttggaag	2100
gccagtttta	acactatgtt	acatttttgt	ttaaagtaag	ttaaaccctta	tataacataa	2160
tgacatttga	tttctggatt	tttcccatgg	ataaaaaatt	aggggggata	aaattaaaat	2220
tg						2222